DOCUMENT RESUME

ED 279 151 FL 016 174

AUTHOR Andersson, Erik; Ostman, Jan-Ola
TITLE Computer Processing of Swedish Syntactic Data. Some Preliminaries and Tentative Results.
PUB DATE 78
NOTE 39p.; In: Andersson, Erik, Ed. Working Papers on Computer Processing of Syntactic Data. Abo, Research Institute of the Abo Akademi Foundation, 1978. p73-108.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Classification; Comparative Analysis; *Computational Linguistics; *Data Processing; Deep Structure; English; Finnish; Form Classes (Languages); Phrase Structure; Research Methodology; Sentence Structure; Structural Analysis (Linguistics); Surface Structure; *Swedish; *Syntax; Uncommonly Taught Languages

ABSTRACT

The encoding of Swedish syntactic information in a study of linguistic data processing is described in detail. The choice of syntactic variables and classification criteria, illustrated with examples, is also discussed. In addition, uses of the coding for contrasting languages, as in this study's comparison of Swedish, Finnish, and English word order, are explored. The coding key consists of 64 variables corresponding to syntactic or semantic properties of a clause. Each clause in the corpus is classified in a subcategory of one of the variables. The variables are divided into four groups: those identifying the clauses, those specifying some syntactic properties of the entire clause or its context, those classifying the clause constituents (verb, subject, object, complement, initial adverbial, adverbial in the central field, non-final adverbial in the end field, and final adverbial in the end field), and those specifying what transformations have been applied to the clause and what constituents they have affected. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED279151

COMPUTER PROCESSING OF SWEDISH SYNTACTIC DATA

Some Preliminaries and Tentative Results

by Erik Andersson and Jan-Ola Ostman

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

N. Enkvist

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL 016 174

BEST COPY AVAILABLE

COMPUTER PROCESSING OF SWEDISH SYNTACTIC DATA
Some Preliminaries and Tentative Results

by Erik Andersson and Jan-Ola Östman

1. Introduction

In the academic year 1976-77, the Text Linguistics Research Group carried out a pilot study of a system for data processing of Swedish, Finnish and English grammar. This paper gives a brief survey of the methodological preliminaries and some tentative results from this study. The aim was to shed some light on the textual factors governing word order, especially in those cases where these factors interact with intrasentential factors and several alternative word orders are available. The basic data were processed by means of the computer programme described elsewhere in this volume by Kohonen and Salmela.

The purpose of this paper is to give a fairly detailed account of the variables used in the encoding of the Swedish material, to justify and discuss the choice of syntactic variables and classificatory criteria, and to give some examples of how the system works in practice. We hope that our work will be of interest to people carrying out similar studies or following up the present project. Originally, the project had a contrastive purpose. By using roughly the same coding key for Swedish, Finnish and English sentences, we intended the computer programme to bring out similarities and differences in the word order systems of these languages. This aim has been postponed for the time being, and in this paper we shall only briefly refer to some contrastive applications of the coding key. It should also be noted that the methodology presented in this paper will be relevant for types of syntactic research other than word order studies, as the computer programme described in the paper by Kohonen and Salmela is applicable to the processing of any linguistic data.

The textual material used for the collection of Swedish data consisted of extracts from a fairy-tale story by Astrid Lindgren, Bröderna Lejonhjärta (pp. 201-207, 100 sentences), and from a young

people's novel by Bo Carpelan, Bägen (pp. 5-14, 163 sentences). Since we were interested in how context affects the word order of individual clauses and how contextual factors interact with internal properties of clauses, we chose the clause as our basic coding unit, i.e. each punch card fed into the computer contained one clause. The total number of clauses in the pilot study was about 500. The small amount of data itself indicates the tentative nature of this study, and the results to be presented later in the paper should be taken --at the most--as indications of results which could be found with the use of a larger corpus material. We are also aware of the fact that the results are representative only of a special genre. However, we think that it is advisable to work out a coding system and test it out on quite simple and regular texts.

It was necessary to revise the coding plan a number of times during encoding, and consequently, our data had to be recoded several times. This is another reason why it is expedient to work on a relatively small pilot corpus. In section 2, we shall present a coding key, which is revised on the basis of the results from the pilot study. This section is rather detailed for two reasons. First, we hope that it could serve as a manual for students encoding further material, and secondly, it may give useful information on the kind of difficulties the encoder encounters. In section 3, we give some examples of the results that can be obtained from a study of this kind. However, these results refer to a previous stage of the analysis. It was partly on the basis of these results that we have made some changes in the coding key. These changes will be indicated in section 2. The revised coding key is presented as an appendix.

2. Towards a coding key for computer processing of Swedish

As we already mentioned, each clause constitutes a record in our coding system. The coding key consists of 64 variables, corresponding to syntactic or semantic properties of a clause. For each variable, a number of subcategories are given, and each clause is classified into one of these subcategories. For instance, variable (52) Type of Topicalized Constituent has the following subcategories:

1 Object, 2 Object as quotation, 3 Adverbial normally placed in the central field, 4 Adverbial normally placed in the end field, etc. If a variable is not applicable to a clause--in this example, if no topicalization has applied--the symbol 0 (or blank) is used.

The variables can be divided into four groups. Variables 1 - 6 are used to identify the clauses. Variables 7 - 15 and 50 - 51 are used to specify some syntactic properties of the entire clause or its context. Variables 16 - 49 are used to classify the constituents of the clause, i.e. the verb, the subject, the object, the complement, an initial adverbial, an adverbial in the central field, a non-final adverbial in the end field, a final adverbial in the end field.[1] Some of these constituents may be missing in a clause, but a clause may also contain constituents not covered by the coding key, e.g. a second adverbial in the central field, or a third one in the end field. However, the clauses in our material proved to contain very few "extra" constituents of this kind, and it therefore did not seem necessary to include additional variables for other types of constituents than those mentioned above. Indirect objects might be an exception, if we want to study when an indirect object is used instead of a dative adverbial. The fourth group of variables is related to the second group: variables 52 - 64 specify what transformations have been applied to the clause and what constituents they have affected.

2.1 Identification of clauses

(1) Language of Text. This variable is needed only if texts in different languages are analyzed. It will be necessary in a comparative study of translations: if a clause is given the same clause number as its translation (cf. below), a computer programme can be constructed which will compare the desired properties of the clause and its translation equivalent (cf. Kohonen and Salmela).

(2) Identification Number of Text. This variable is especially useful for stylistic investigations, since we can compare the frequency of a particular phenomenon in one text with that in other

[1] For the definition of central and end fields, see the discussion of nexusfelt and inholdsfelt in Diderichsen 1946.

texts. This can be done by treating the individual texts as sub-corpuses of their own (using the text numbers as the "filter variables") or by cross-tabulating variable (2) against the variable under investigation.

(3-5) Number of Sentence in Text. The sentences in each text are numbered consecutively from 001 to 999. If longer texts are used, a fourth (and possibly a fifth) column has to be added.

(6) Number of Clause in Sentence. In contrastive studies, it is practical to give each clause its identification number in two steps, i.e., via a sentence number. This is because the division into sentences is usually less controversial than the division into clauses. It is quite likely that a translation will contain correspondents to all the sentences in the original text, and these sentences can be matched quite easily. At clause level, however, it often happens, for example, that two clauses in one language correspond to one clause in the other language. Similarly, it is sometimes hard to determine whether a construction should be regarded as a clause and thus receive a card of its own, or whether it should be regarded as a non-clausal constituent (e.g. objects with infinitives and various incomplete clauses). If the clause number is given via a sentence number, a re-coding at clause level does not necessitate recoding of all the sentence numbers in the text.

In contrastive studies, it is advisable to match the clause numbers only in so far as the matched clauses really are comparable. A very "free" translation should not be given the same clause number as the original clause. If the Swedish original contains three clauses numbered 1, 2, and 4, the Finnish translation might contain four clauses numbered 1, 3, 5 and 6. Such gaps in the numbering are much easier to arrange if clause numbers are given via sentence numbers.

Cleft constructions of the type Det var Kalle som gjorde det 'It was Kalle who did it' were regarded as single clauses. The reason was twofold: these constructions have partly the same function as word order permutation in a single clause (topicalization, passive), and the cleft sentence in Swedish corresponds to a single clause in Finnish.

Coordinated predicate phrases were regarded as separate clauses.

One reason was that predicate phrases contain the kind of functional elements classified by our coding key (object, complement, adverbial); another was that it is often impossible to distinguish coordinated clauses from coordinated predicate phrases in Finnish, where the subject can often be deleted. However, coordinated infinitives were not regarded as separate clauses, and here we had to make a decision always to classify the functional elements of the latter infinitive later on in the coding.

2.2 Clausal variables

(7) Number of Clauses in Entire Sentence. This variable is relevant especially in stylistic investigations, since the complexity of sentences--in terms of number of clauses per sentence--varies between texts. But it is also possible that this variable may affect the inner structure of the constituent clauses--complex sentences might favour certain transformations and word order tendencies, etc.

(8) Matrix Address of Dependent Clause. This variable specifies the clause which is hierarchically above the one being coded. Variables (3-5) specify the sentence to which both these clauses belong. Variable (8) was added to the coding key after the preliminary analyses, and the present computer programme cannot make much use of this variable. But when the programme has the necessary module (cf. Kohonen & Salmela), it can be used to compare codings on two different cards, i.e., in comparisons between a dependent clause and its matrix clause.

(9) Status of Clause in Sentence. All clauses were classified as either main clauses or dependent clauses. The dependent clause was here defined as a clause functioning as a modifier in a matrix clause. No attention was paid to the inner structure of the clause, e.g. the difference between main clause and dependent clause word order. The main clauses were subdivided into non-cocordinated and coordinated clauses. The latter were further classified according to their position in the coordination, whether initial, medial, or final. In the provisional coding key, non-coordinated clauses and initial coordinated clauses formed a single category.

The dependent clauses were classified as occurring before,

inside or after the rest of their matrix clause. If two clauses were standing last in their matrix clause, they were both classified as final, i.e. a clause is regarded as matrix medial only if it is surrounded by non-clausal constituents. The dependent clauses in end-position were further divided into two subcategories: sentence-final, and not sentence-final clauses.

(10) Clausal Construction. All clauses were classified according to their inner structure as simple (containing no dependent clauses) or complex (containing dependent clauses). Furthermore, the clauses were classified as non-coordinated and coordinated, and, if co-ordinated, as initial, medial, or final in the coordination. To a certain extent this repeats the information already given by variable (9). The difference is, however, that these figures concern all the clauses in the material, while variable (9) classified main clauses only as to coordination. It would then have been possible to remove the coordination classification from variable (9) without losing too much information, but it might be convenient to have direct access to the figures for main clauses only, without having to filter out the dependent ones.

(11) Clause Type. A basic distinction was made between main clauses and dependent clauses, with a main clause now being defined as a clause displaying main clause word order and a dependent clause as a clause displaying dependent clause word order. This holds if the distinction in word order is observable; otherwise typographical criteria or the embedding criterion (cf. variable (9)) was used here, too. Main clauses were further subdivided into main clauses introduced by a coordinator (och 'and', men 'but', för 'since', etc.), embedded main clauses and other main clauses. Originally, the last category included only embedded direct quotations, but we suggest that other embedded clauses with main clause word order should be included, too, such as the following case:

Ser du en taxi, så ta den.
'If you see a cab, take it.'

The dependent clauses were subdivided into interrogative clauses, att-clauses without preposition, att-clauses preceded by a preposition, other adverbial clauses, relative clauses of three different types (those modifying the subject, those modifying the object, and

those modifying other constituents), comparative clauses and clausal contractions (e.g. object with infinitive). Originally, interrogative clauses did not form a separate category and there was only one category of relative clauses.

Att-clauses where att ('that') has been deleted were still classified as att-clauses. All dependent clauses introduced by a question word were regarded as indirect interrogative clauses:

Jag kunde inte tala om vad filmerna egentligen handlade om (02,60:6)
'I could not tell what the films really were about'[1]

Jag kände hur jag flög fram (02,124:3)
'I felt how I flew forward'

That is, the matrix clause need not contain an interrogative verb, not even a verbum dicendi. Free relative clauses of the following type, however, were treated as relative clauses:

Jag dansar var jag vill
'I dance where I want to'

(12) Surface Mood of Clause. This variable classified clauses into affirmative, interrogative, imperative and exclamatory clauses. Strictly formal criteria, such as the use of a question word or subject-verb inversion in questions, the tenseless imperative form in imperatives, etc. were decisive. Exclamatory clauses form a heterogeneous category including greetings and thanksgivings. The manual will have to be diversified on this point. One could also suggest a classification of clauses according to functional mood, i.e. according to speech acts (cf. Hakulinen & Karlsson, forthcoming). In that case, a new variable would have to be introduced.

(13) Clause Structure. The clauses were subclassified into active transitive and active intransitive clauses, passive clauses with an overt subject, clauses containing a formal subject, and fragmentary clauses. Clauses containing a formal subject were further subdivided into existential clauses (with an existential subject in object position), subjectless passives of the type Det dansades 'There was some dancing', and other clauses with a formal subject, e.g. Det regnar 'It is raining'. Examples like Det smiddes vapen 'Weapons were forged' were classified as existential clauses, like

[1] The figures in parentheses after the examples refer to text number, sentence number, and clause number, in that order.

the type Det finns svamp i skogen 'There are mushrooms in the forest', although it is not totally clear that vapen should be regarded as a subject -- it could also be analyzed as an object. In that case the clause would be analyzed as a passive construction with a formal subject only.

Passives consisting of bli + participle were coded as active sentences. However, this decision is open to question.

(14) Surface Word Order of Clause. In the provisional coding key over 50 different surface word order patterns were distinguished, and consequently less frequent categories were exemplified by no more than a couple of clauses. If the statistics are to be meaningful, the size of the corpus should be drastically enlarged, so that a fair number of clauses will fall under each category. Even so, it will be hard to handle so many categories. We have therefore reduced the number of surface patterns by studying only the three main constituents of the sentence: the subject, the verb, and one complement (in a wide sense). The alternatives will then be the six possible permutations of these three constituents, and also the patterns SV, VS, and V. However, there are still many unsolved problems in this system. If there are several complements, which one should be chosen as the basis of the classification? Is V the finite verb or the main verb?

Elliptical constructions also pose a problem. In this pilot study the missing elements were indicated "as if they were there", in the position where they would be placed most conveniently.

In clauses with a formal and an existential subject the existential subject was regarded as S, and the formal subject was ignored. In clauses lacking a real subject the formal subject was treated as S, but it could be ignored here, too,

Cleft constructions were classified according to the word order of the higher sentence in the construction. E.g. Det var Johan som jag såg 'It was John who I saw' was coded as SVX. However, the coding would have been more consistent if the sentence had been coded as XSV, because the higher sentence Det var ... was disregarded when clause numbers (variable (6)) were assigned. Also, we are interested here in the thematic ordering of the main constituents of the clause, not the ordering of such formal elements as det and var, which is

fixed by quite simple and well-known rules.

(15) Verb Complements. This variable is intended to show what type of constituent realizes the X mentioned in the preceding variable. X was classified as an object, a true complement, or an adverbial. The variable was added later to the coding key.

The provisional coding key also contained a couple of variables which were subsequently removed. One variable classified clauses according to type of negation: normal negation by means of contractions like ingen 'nobody', or by means of quasi-negative adverbials like knappast 'hardly'. This variable might perhaps be preserved if the aim is to contrast Swedish and Finnish. Another variable in the provisional coding key indicated the occurrence of interjections before, after, or in the middle of a clause. Although it is possible that an interjection or a vocative expression might trigger a topicalization or affect word order in some other way, this variable has been excluded from the coding key in its present form.

2.3 Constituent variables

2.3.1 The verb

(16) Verb Structure. This variable describes what in traditional grammar has been called the predicate proper. Although the verbs of a sentence need not always form a constituent, we have grouped them together for practical reasons. The variable has four different dichotomies: a) simple vs. complex, i.e. forms containing a modal verb or another auxiliary verb, e.g. simmar - kan simma, har simmat; b) simple vs. reflexive, e.g. öva - öva sig; c) simple vs. phrasal, i.e. verbs consisting of several words, e.g. slå - slå till; d) simple verb vs. verb chain, i.e. a sequence of several main verbs, e.g. springa - börja springa. When these four dichotomies are combined, a system containing $2^4 = 16$ subcategories is obtained.

The phrasal verbs typically contain a verbal particle. However, we have also included idiomatic coordinations of verbs in this category, e.g:

Han gick och gick. 'He walked and walked.'
Han satt och skrev. 'He sat writing.'

These coordinations do not contain two independent, equal verbs. The borderline between these constructions and true coordinations is, however, blurred. The alternative is to treat the constructions as forming two separate clauses, since in our system each predicate phrase is taken to constitute a separate clause.

Special problems are also raised by verbal expressions containing nouns, as hugga huvudet av 'decapitate'. The particular aim of one's study will have to determine whether such expressions should be treated as phrasal verbs, or as combinations of a verb and an object.

Two further variables classifying the verb were used in the provisional coding plan. One variable specified the semantic properties of the verb, using verbs like känna 'feel', verbs of saying, verbs denoting actions and happenings, stative verbs, and verbs denoting change of state. The intention was to test the usefulness of this classification as compared to another aspectual classification into stable states, temporary states, irresultative events, durative events, and momentary events (cf. Andersson 1977). Since the coding proved quite cumbersome, we have decided to postpone the study of these variables, thereby giving space to other, more textually relevant variables.

2.3.2 The subject

(17) Subject Structure. This variable specifies the internal complexity of the subject. The subcategories were coordinated subject; subject containing subordinate clause; clausal subject; infinitival subject; subject followed by a modifier (not clausal); subject preceded by a participial modifier; subject preceded by some other modifier; one-word subject, possibly preceded by a separate article; formal subject; and deleted subject (usually gapped). Since some subjects fit several of the categories, we decided to impose a hierarchical ordering between the subcategories, so that a subject was classified as a coordination rather than as a subject containing a subordinate clause, as a subject containing a postmodifier rather than as one containing a premodifier, etc. In the revised coding plan, the subcategories are presented in this hierarchical order. The same applies to several other variables.

When a clause contains both a formal subject and an existential subject, this variable, like the other subject variables, classifies the existential subject. The subcategory formal subject will then be applied only to clauses containing no other subject than a formal subject, as Det regnar 'It is raining'. Figures for all clauses containing a formal subject will anyway to some extent be obtained from variable (13).

(18) Subject Form. This variable is mainly concerned with the definiteness marking of the subject. Indefinite subjects are those with the indefinite article, with (or consisting of) an indefinite quantifier, and without either article or quantifier. Definite subjects are those with the definite article, with (or consisting of) a definite quantifier, with a genitive attribute, or proper names (including words like mamma 'mummy' and pappa 'daddy') or definite pronouns. Clauses and infinitives etc. fall into the category of "irrelevant". Examples of indefinite quantifiers are några, fler, ingen(ting), vem som helst, and the numerals, and examples of definite quantifiers are alla and de flesta. Sometimes the context will give hints as to the choice between these categories.

Coordinated subjects are problematic, since our policy was to let the first part of the coordination determine the classification, while the other parts are ignored.

(19) Givenness of Subject, and (20) Type of Coreference of Subject. These two variables are discussed in section 2.5, since the criteria are the same for all nominal constituents in the sentence. Variable (21) Number of Words in the Subject Phrase, shows the length of the subject.

2.3.3 The object

(22) Object Structure. The structure of the object is comparable to that of the subject. An example of a formal object is det in jag har det bra 'I am having a good time'. The category 'Gapped subject' is parallelled by the category 'Deleted object', which covers both ellipsis under identity and sentences with an implied object such as Han åt 'He was eating'.

(23) Object Form. The same as for the subject. (24) Givenness of Object. (25) Type of Coreference of Object. (26) Number of Words in Object Phrase. Indirect objects cannot be treated properly by this

coding key. One alternative would be to encode them as complements, another to encode them as adverbials, since they can be replaced by dative adverbials. Cf. section 2.4.

2.3.4 The complement (Sw. predikativ)

(27) Complement Structure. Nominal complements were classified into roughly the same categories as subjects (the categories 'Formal subject' and 'Gapped subject' have no correspondents). Adjectival complements were classified as coordinations, adjectives modified by a clause, e.g. Han var så lat att . . . 'He was so lazy that . . .', adjectives with other modifiers, e.g. mycket klok 'very wise', and simple adjectives.

The provisional coding plan included a variable describing the form of the complement (cf. the subject and the object), but this variable was regarded as superfluous for the present purpose.

(28) Complement Type. Complements were classified as optional or obligatory, and as specifying the subject or the object.
(29) Number of Words in Complement.

2.3.5 The adverbials

In each clause, the coding key can handle a maximum of four adverbials. That is, we have identified four positions for adverbials in the clause and ignore adverbials in any other position. The existence of adverbials in other positions will generally be indicated under the variable measuring the total number of adverbials in the clause. -- The four types of adverbials are the following:
A. Initial adverbials, which according to Diderichsen (1946) are placed in the fundament field, i.e. topicalized adverbials, potentially preceded only by coordinating or subcoordinating conjunctions and dislocated elements.
B. Central field adverbials, which in subordinate clauses precede the finite verb, and in main clauses follow the finite verb and the subject but precede the infinite verb and/or the object.
C - D. End-field adverbials, which follow even infinite verb forms and objects. The coding plan can take care of two adverbials of this

category: the last but one and the last adverbial. The latter can be followed only by dislocated elements.

If a main clause contains no infinite verb or object, central field adverbials cannot be distinguished from end field adverbials. In such cases, a test based on the intuitions of the coder has been carried out: the clause has been changed into a subordinated clause, or an infinite verb form has been added. If these tests do not favour either position, the adverbial has been entered as an end-field adverbial.

It should also be noted that two adverbials can be joined to form a compound adverbial. They are then classified as one adverbial; the criterion is whether they can be topicalized together. -- The variables given below are the same for all four types of adverbials.

(30, 35, 40, 45) Adverbial Structure. The following classifications were used: one-word adverbials; phrases not introduced by a preposition; phrases introduced by a preposition; infinitives (possibly preceded by a preposition); clausal contractions of some other type; full clauses, possibly preceded by a preposition; noun phrases or prepositional phrases containing a clause. Articles were not counted as words.

The clausal contractions are mainly reduced comparative clauses like: Han gråter mer än Kalle 'He weeps more than Kalle'. The following are examples of an adverbial infinitive without a preposition: Och en enda liten flamma räcker att förlama eller döda . . . 'And one single little flame . . . is enough to paralyse or to kill . . .' (01,51:1) Han har gett oss en mur att skydda oss med 'He has given us a wall to protect ourselves with' (01,54:2).

(31, 36, 41, 46) Adverbial Type. Negations, commentatory adverbials, like också 'also' and bara 'only' formed special subcategories. Special clausal adverbials are often distinguished by their ability to be combined with a noun phrase in the clause, as in Också Kalle var där 'Kalle also was there'. Such combinations are of course not treated under these variables, which classify elements in adverbial positions (cf. variable (50)). Another subcategory under this variable is 'frame adverbials', i.e. place and time adverbials which present a setting for the event of the clause. Other adverbials are classified as optional adverbials; and obligatory ones (mainly

valency adverbials).

(32, 37, 42, 47) Semantics of the Adverbial. This variable classifies the adverbials according to the familiar categories of time, place, manner, degree, measure, purpose, goal, condition, cause, instrumental, and agent (in passive constructions). A special category called "irrelevant" is used for other types, such as negations and special clausal adverbials, which had already been coded under the preceding variable.

(50) Special Clausal Adverbial inside a Functional Element. The classification distinguishes between adverbials in the subject, object, complement, and another adverbial. The variable is perhaps not essential to the coding key.

(51) Number of Adverbials in the Clause. Only adverbials modifying the predicate verb of the clause are counted, not adverbials inside a functional element in the clause.

2.4 Textual variables

(19, 24, 33, 38, 43, 48) Givenness. The classification is the same for subjects, objects, and adverbials, although the category 'irrelevant' will normally be used for adverbials not consisting of a noun phrase or a prepositional phrase.

This variable indicates to what extent the REFERENT of a noun phrase has been previously given or introduced into the discourse universe of the text. The basic classification is between noun phrases referring to a new referent and those referring to an old referent. The category 'given' is further subdivided into several classes. The referent can be textually given or pragmatically given, i.e. given in the speech situation or through the common knowledge of the participants in the speech situation. Pragmatically given elements formed a single category in the coding key, but were of three kinds: performatively known noun phrases, referring to some of the participants in the speech situation, generic noun phrases, usually referring to the whole of a group of individuals, and noun phrases referring to a generally known referent, e.g. solen, 'the sun'. Textually given elements were classified as mentioned in the preceding sentence (or even in the same sentence), or as previously

mentioned. A third category might be useful, 'recently mentioned', and could cover referents mentioned in the same passage, or referents which have not disappeared from the scene since they were mentioned.

Indexically given referents formed an intermediate category between textually and pragmatically given noun phrases. These referents normally occur together with a referent which is mentioned before (e.g. driving - the steering wheel) or form a part of an earlier mentioned referent, (car - steering wheel), or belong to a group of elements which is given as a whole, but does not contain any previously distinguished elements (flowers - a tulip; boys - one of the boys). The latter category should probably form a subcategory of its own, since it is related to the category of new referents.

The givenness of clauses and infinitives was generally regarded as "irrelevant", since judgements otherwise would have been extremely hard.

In coordinations of NPs the functional element as a whole is indicated as given if ONE of the coordinated elements is given. This also applies, mutatis mutandis, to the variables accounting for coreference of NPs.

(20, 25, 34, 39, 44, 49) Type of Coreference of the Noun Phrase. This variable indicates in what way an element is textually bound to the rest of the text. It mainly indicates the type of relation between a given referent and its correlate. However, the variable or a corresponding one could also be used to indicate similar relations between elements inside the noun phrase and the context. That is, one variable indicates the relationship between the referent of the noun phrase and the context, and another one indicates the relation between the noun phrase expression and the context.

The subcategories indicate the type of linking used: by means of a pronoun (anaphoric, cataphoric or exophoric), or by means of the repetition of a lexical noun, or by a paraphrase. The category "irrelevant" includes clauses, infinitives, and textually free elements, but the latter category could also constitute a subcategory of its own.

2.5 Transformations

The aim of the project was to study the motivations for the use of a particular word order. Our hypothesis was that a transformation is

often applied to a clause in order to make the syntactically "basic" word order fit into the context better. It was thus natural to work within the theory of transformational grammar, and to classify clauses according to what transformations they have undergone. We have consequently been interested in optional transformations, mainly movement transformations and deletions. Our goal has been to reduce the optionality of these transformations by studying the factors which trigger them off. However, we have concentrated on transformations quite close to the surface, which makes it possible to avoid taking a standpoint in the controversy between Interpretive and Generative Semantics.

2.5.1 Movement transformations

(52) Type of Topicalized Constituent. Topicalization is defined as the movement of a constituent to initial position without insertion of a pause or a comma between the moved constituent and the rest of the clause. The topicalized constituent can therefore be preceded by coordinating or subordinating conjunctions and by dislocated elements. Subjects were not generally indicated as topicalized, but a special category of subject-initial clauses should probably be introduced under this variable. Topicalized objects are of two kinds: quotations and others. The motivations for the movement is probably different in these two cases. Adverbials are classified as central field adverbials and as end field adverbials, according to the coder's intuition where the adverbial most naturally could be placed if it had not been topicalized. Topicalized complements and verb phrases are less controversial subcategories. An example of a further subcategory, 'topicalized subject in an existential sentence' is the following: Sådana finns det här 'There are ones like that here'.

The same variable can accommodate other transformations, too, if desired. The condition is of course that the transformation is complementary to the topicalization transformation. Yes-no questions, where the finite verb is moved to initial position, can therefore be recorded under a special subcategory. The transformation movement of a relative pronoun is also complementary to topicalization, and could even use the same subcategories. The filtering device of the computer

programme could be used to distinguish between topicalization and relative pronoun movement: if the computer operates on a subcorpus of main clauses only, it will indicate topicalizations. This requires that the category main clause be defined on the basis of word order rather than on the basis of embedding. A similar solution is not possible for question-word movement, since this transformation applies both in main and in subordinate clauses. This transformation therefore requires subcategories of its own.

(53) Subject Postponed to the End-Field. A normal subject-verb inversion is not recorded under this variable. What is required is that the subject be moved to the end field, i.e. some constituent has to intervene between the finite verb and the subject, preferably not a short central field adverbial. The subcategories are moved subject clause, moved subject infinitive, NP subject in existential clause (containing the formal subject det), and other moved NP subject.

It would also be interesting to know when extraposition of the subject has not taken place, although possible. Special categories could therefore be introduced for non-extraposed subject clauses and subject infinitives. Similarly, sentences in which the existential construction is allowed by the verb and the definiteness of the subject, but not chosen by the author, could be recorded under this variable.

(54) Extraposition and Quantifier Movement. This variable accounts for extrapositions of parts of the main functional elements in the clause. The extraposed element is normally an att-clause or a relative clause:

Den omständigheten förvånade mig att Pelle hade varit hemma.
'That circumstance astonished me that Pelle had been at home'

Jag såg en man igår som hade rött hår.
'I saw a man yesterday who had red hair.'

The subcategories are extraposition from the subject; the object; the complement; and from an adverbial.

Quantifier movement forms a separate category, e.g. Pojkarna var alla där 'The boys were all there': a quantifier is moved from a noun phrase into adverbial position. A further minor category is topicalization of part of the object: Bilar hade vi fyra 'Cars, we had four of', where the quantifier is left in the normal object position.

Of course, the transformations 'extraposition of a modifying clause' and 'quantifier movement' may both apply to the same sentence. If so, one of the transformations will be left unrecorded. However, these transformations seem to be so infrequent that such cases will be extremely rare and can therefore be excluded from computer processing, to be dealt with manually. Otherwise the variable will have to be split into two.

A special problem is created by constructions of the type V_1 Adv och V_2, where V_1 och V_2 is analyzed as a phrasal verb. If the adverb intuitively belongs to both verbs, V_2 can be analyzed as an extraposed verb. Here an alternative analysis has been applied: the adverb is regarded as having been moved into the phrasal verb, or rather, the first verb in the phrasal verb has been moved in front of the adverbial. This is an automatic transformation in main clauses (finite verb movement]. It has not been taken into the coding key as a special transformation.

It is not always necessary for the extraposed element to be separated from its mother constituent by some other element. The separating element can also be another modifier of the same constituent, which is normally placed after the extraposed one. In the following example också is the separating element: Det fanns en vik också med låga buskiga tallar (02,3:1) 'There was a bay, too, with low, bush-like pine-trees'.

An example of an extraposition of a functional element would be: Jag skulle ha det kvar, alltid. 'I would have it in me, forever' (02,140:2).

(55) Type of Cleft Construction. Since Swedish cleft constructions correspond to Finnish word order permutations in a single clause, it was natural to regard the cleft construction as consisting of one clause rather than two. Support for this analysis can also be found in the fact that the cleft construction seems to be a rather late transformation, preserving e.g. the case marking of the single clause: cf. Det var jag som kom. 'It was I who came' and Det var mig som du såg. 'It was me that you saw'. The characteristic of the cleft construction is the presupposed som-clause. Sentences in which the som-clause is not presupposed were therefore not analyzed as cleft constructions, e.g. Det var mamma som ropade från köket (02,90) 'It was Mummy, who shouted from the kitchen'. Here it is not clear from the context that somebody was shouting from the kitchen. The sentence is

therefore analyzed as an existential clause containing a non-defining relative clause. However, the borderline between these two constructions is very unclear.

The clefted element was classified as subject, object, complement, or adverbial. An additional subcategory was used to indicate a pseudo-cleft construction, e.g., Vad jag tänkte på var detta 'What I was thinking of was this'. This was possible because the cleft transformation seems almost never to be applied to a pseudo-cleft sentence.

(56) Left Dislocation. The basic criterion for dislocation is that a functional element be moved outside the clause, so as to form a tone unit of its own. This element is represented inside the clause by a pronoun copy, which is often topicalized, e.g. På gården, där hade vi lekt 'In the yard, there we had been playing'. The subcategories classified the dislocated elements according to their function in the clause. It is also possible to distinguish between topicalized and non-topicalized objects, adverbials etc.

(57) Right Dislocation. The classification is the same as for the preceding variable. The dislocated element is placed last in the clause, e.g., Där hade vi lekt, på gården.

(58) Raising. The subcategories were subject raising, e.g. Jag såg honom komma 'I saw him come', generic object raising (or Tough-movement), e.g., Fiolen är svår att spela på 'The violin is hard to play', and non-generic object raising, e.g., Fiolen är svår för mig att spela på 'The violin is hard for me to play'. Object raising is really applicable to all non-subjects, and different subcategories could be used for objects and adverbials.

Originally clauses in which raising could have applied but had not been used were also recorded. This was later dropped as too cumbersome, but would have provided interesting data.

(59) Passive. This variable could be used for several purposes. One possibility is to indicate whether an agent adverbial is present in the clause or not, and where it is placed. This information is coded elsewhere in the coding-plan, but it is convenient to be able to pick out information from one single variable rather than collecting it from several variables. Another possibility is to indicate whether the passive is a bli-passive or an -s-passive. A third possibility is to record sentences constructed with the subject man 'one' under a

special subcategory.

2.5.2 Elliptical transformations

(60) Verb Phrase Deletion. This variable is used to indicate that a verb phrase has been deleted under identity with a verb phrase in the context (cf. Andersson 1976). The variable is not very central to the coding key. The subcategories could be used to indicate whether the first or the second verb phrase is deleted, and whether the deleted verb phrase is not contained in the remaining one (outer deletion), or, is strictly speaking a part of its own correlate (inner verb phrase deletion). We then get the following examples:

Du får simma om du vill /. 'You may swim if you want to'	-forward, outer
Du får vad du vill /. 'You get what you want'	-forward, inner
Om du vill / får du simma. 'If you want to, you may swim'	-backward, outer
Vad du vill / får du. What you want, you will get'	-backward, inner

(61) Ellipsis of Functional Elements (Gapping). An ellipsis was recorded under this variable only if one or several elements were deleted under identity with the corresponding elements in the context. Pragmatic ellipses, such as subject deletion in imperative clauses, were not indicated under this variable. In general, the elliptical sentence should contain at least two different functional elements of the sentence, so as to distinguish the ellipsis from a normal coordination, which has not been recorded at all here. The only exception to this rule is coordinated predicate phrases, which have been indicated as two (or several) different clauses. They are coded for subject ellipsis. The other categories are used for true gapping, and often involve deletion of the verb. The different subcategories indicate what functional elements have been deleted.

(62) Ellipsis of Part of the Verb. Under this heading were gathered some less interesting deletions, which do not seem to have a clear textual motivation: deletion of ha 'have' in dependent clauses

or after a modal verb, deletion of göra 'do' or fara 'go'. The variable could also be used to indicate that the predicate contains a coordination of infinite verb forms.

(63) Ellipsis of the Head of a Noun Phrase. The subcategories of this variable are used to indicate whether the ellipsis took place inside a coordination of noun phrases or between two other noun phrases, and whether it is the head of the first or the second noun phrase which has been deleted.

Den snälla/och den dumma gossen 'The good and the bad boy'	coord., backward
Den snälla gossen och den dumma / 'The good boy and the bad one'	coord., forward
Den sjuka / såg den friska pojken 'The sick one saw the healthy boy'	other, backward
Den sjuka pojken såg den friska / 'The sick boy saw the healthy one'	other, forward

(64) Number of Transformations in the Clause. Only the transformations listed in the coding key (variables 52-63) were counted.

3. Some preliminary results and hypotheses for future research

The quantitative results yielded by the present system can be utilized in at least three ways. Firstly, the raw frequencies of the subcategories of a single variable can be of interest. Secondly, two variables can be cross-tabulated, which makes it possible to examine their interdependence. Thirdly, the data can be used to compare two languages, of two or several texts or genres in the same language. In this section, some examples of these types of results will be given. However, these results should be regarded as hypotheses to be tested by further research, since the pilot study was limited to a very small material.

3.1 Frequencies of subcategories

For variable (10), Clausal Construction, the pilot study gave the following results:

	not co-ordinated	first part in coordination	middle part in coord.	last part in coord.	total
simple	168	91	42	76	377
complex	63	15	2	28	108
total	231	106	44	104	485

Out of 485 clauses, 108 or 22.3 % were complex, i.e. contained one or more dependent clauses, and 254 clauses, that is, 52.4 % were coordinated. If we compare the different positions in the coordination, we find that 15 out of 106 first parts or 14.2 % were complex, while 28 out of 104 last parts in the coordination or 26.9 % were complex.[1] That is, last parts in coordinations were generally more complex than first parts. This tallies with the principle of end weight. Complex middle parts in coordinations were very rare (4.5 %); however, here the comparison is somewhat misleading, since the middle parts were compared to first and last parts in general, not only to first and last parts in a coordination with a middle part. Out of the sentences which were not coordinated 63, or 27.7 %, were complex, i.e. the figure was comparable to the one for last parts in coordinations.

Variables (18), Subject Form, and (25), Object Form, classify the subject and the object according to their marking for definiteness. If we add together the figures for the various indefiniteness markings in order to get larger categories, we obtain the following diagram:

	missing or deleted	indefinite	definite	irrelevant	total
subject	105	35	327	18	485
object	296	38	99	52	485

The high figure for objects with irrelevant definiteness reflects the fact that objects are sentential much more often than subjects. Out of the 380 subjects, 35, or 9.2 %, were indefinite, and 327, or 83.4 %, were definite. The corresponding figures for the 189 objects are 38,

[1] Here we find a small coding (or punching) mistake. The number of the first parts in coordinations should of course equal the number of the last parts.

or 20.1 %, indefinite ones, and 99, or 52.4 %, definite ones. That is, subjects have a much stronger tendency towards definiteness than objects. This tallies with the text linguistic principle whereby given elements, which often function as topics, are placed in the beginning of the clause and high up in the syntactic structure. Similar figures can be obtained from the variables classifying the givenness of the subject and the object referent, (19) and (24).

3.2 Cross-tabulations

3.2.1 Topicalization and subject length [1]

A cross-tabulation was carried out on a subcorpus of the main declarative clauses, i.e., on those marked both for one of the subcategories 1, 2, or 3 of variable (11) and simultaneously for subcategory 1 of variable (12). The two variables cross-tabulated were (17), Subject Structure, and (52), Type of Topicalized Constituent. However, since the topicalizations in the material were relatively few, the figures in subcategories (1-7) were combined. We then obtain an overall figure for topicalization, which can be compared with the figure for sentences where no topicalization applied (subcategory 0).

structure of subject	topicalization	no topicalization	total
deleted	2	86	88
formal det	5	11	16
one word	90	121	211
with preposed attribute	9	11	20
with postposed attribute	2	12	14
clause or infinite	0	6	6
coordination	3	7	10
total	111	254	365

1 Within a similar framework, topicalizations in Swedish have been analysed recently by Marjatta Lahti in her graduate thesis (1978), using the present data processing system.

In the clauses with no topicalization, the subject almost always occupied the initial position. After topicalization, the subject was placed in the central field, immediately after the finite verb. That is, as a consequence of topicalization, the subject was placed inside the clause, often with other functional elements on both sides. The only exception to this rule were existential clauses, in which the subject was placed in the object position (i.e., after the main verb) in both cases. Existential clauses should therefore have been excluded from this table, but their frequencies were so low that they hardly influence the statistics.

We can easily see that there is a strong tendency to favour topicalization when the subject consists of one word or has only preposed attributes. If the subject is gapped, or consists of a clause or an infinitive, or even consists of a postposed modifier, topicalization takes place much more seldom. The rule could be that the central field should preferably contain short elements. We have a centripetal tendency in Swedish: the principle of end weight works in both directions.

3.2.2 Topicalization and the form of the subject

If variable (17), Subject Structure, in the above cross-tabulation is changed to variable (18), Subject Form, the following results are obtained:

form of the subject	topicalization	no topicalization	total
missing	2	86	88
indefinite	12	12	24
definite			
- proper name	19	20	39
- defin. art.	13	35	48
- genit. attr.	1	8	9
- definite quantifier	59	75	134
- pronoun	2	5	7
irrelevant	3	13	16
total	111	254	365

The table shows that topicalization is relatively frequent when the subject is indefinite or a proper name, less frequent when it is a pronoun, and rather infrequent when it is a definite noun phrase of some other type. This is in accordance with the rule that new referents tend not to be placed in initial position.

3.2.3 Topicalization and clausal structure

If the subject variable in the preceding cross-tabulations is changed to variable (13), Clause Structure, the following results are obtained:

	topicalization	no topicalization	total
active transitive	61	147	208
active intransitive	43	133	176
passive	0	4	4
existential	2	10	12
other clauses with a formal subject	5	11	16
fragments	0	10	10
total	111	315	366

As we can see from the figures, topicalization is somewhat more frequent in transitive sentences (29.3 %) than in intransitive sentences (24.4 %). It is possible that this difference is due to the fact that transitive clauses contain more constituents which can be topicalized. A way to test this hypothesis would be to cross-tabulate variable (52), Type of Topicalized Constituent, and variable (51), Number of Adverbials in the Clause. However, it should be noted that in active transitive clauses, 29 quotation objects are topicalized. If clauses containing quotations are disregarded, the difference between transitive and intransitive sentences might be smaller or non-existent.

3.2.4 Givenness of final adverbial in main and subordinate clauses

Variable (48), Givenness of Final Adverbial, was cross-tabulated against variable (11), Clause Type. We have here added the figures for all main clause types, and similarly for all dependent clause types, in order to obtain larger and more reliable figures:

	main clauses	dependent clauses	total
no final adverbial	216	49	265
just mentioned	20	7	27
earlier mentioned	16	9	25
indexal givenness	46	6	52
generally implied	7	1	8
new referent	29	6	35
irrelevant	49	8	57
known from the speech situation	10	2	12
total	393	88	481

We can see from the table that dependent clauses contain a final adverbial which has been mentioned before about twice as often as main clauses do (10.2 % of the dependent clauses and 4.1 % of the main clauses). This seems to support the hypothesis that dependent clauses contain more given information than main clauses. The tendency to put a new referent last in the clause seems to be stronger in main clauses. One of the reasons for this might be that dependent clauses often are presupposed, referring to previously known facts.

3.3 Comparison between Swedish and Finnish

As indicated in the introduction, one of the basic aims with our project was to create a common ground for comparing the occurrence of textual and syntactic phenomena in different languages. In fact, we even used the same texts as material for the Finnish and Swedish pilot

studies. (Both texts for the Finnish study were translations of the texts used for the Swedish coding; cf. Hakulinen & Karlsson, forthcoming.)

The task of finding a level at which the different languages would be fully comparable is, of course, a difficult one. Each language is by itself a phenomenon built up of interacting systems. And it will not do to try to isolate a particular factor from one system and compare it with a similar factor in another language without taking into account the fact that this factor in turn also belongs to a system; and though we can compare systems in different languages, they rarely overlap in all their details.

With this in mind we shall here give just one example of what kinds of information similar coding keys for different languages can give. (52) Type of Topicalized Constituent. Different types of functional elements seem to be topicalized with different frequencies in Swedish and Finnish:

topicalized constituent	Swedish	Finnish
no topicalization	356	410
quotations	31	
other objects	18	84
adverbial		92
- from end-field	51	
- from central field	21	
complement	6	
existential subject	1	
verb phrase	1	
total	485	602

As can be seen from the figures, the object seems to be topicalized much more frequently in Finnish than in Swedish. (This is true even if we add the frequencies of topicalized quotations to the Swedish figure.) One of the most obvious reasons for this is that Finnish--as opposed to Swedish--is a highly synthetic language, and can make use of inflectional devices to indicate the grammatical functions of NPs in a clause. Another difference that could be noted was that temporal adverbials were topicalized remarkably often in Swedish. In Finnish, place adverbials were topicalized with the same frequency as temporal ones.

APPENDIX 1.

Revised version of the TLRG coding key for data processing of Swedish

In this section we wish to give a summary of the variables discussed in § 2, and their use. The numbers on the left refer to the columns on the punching card, and are indicated here for ease of reference. Needless to say, the numbering used below is arbitrary, and can be varied. The different values or subcategories of the variables are also numbered (1 - 9, followed by A - Z; thus A equals subclass number 10, etc.) For variables where the subclasses overlap, we have tried to indicate what we feel to be the "hierarchical order" of the subclasses by giving the 'strongest' subclass number 1, etc.

COLUMN	VARIABLE	VALUES
Identification of Clauses.		
1	Language of Text	1 Finnish
		2 Swedish
2	Identification Number of Text	
3 - 5	Number of Sentence in Text	
6	Number of Clause in Sentence	
Clausal Variables.		
7	Number of Clauses in Entire Sentence	
8	Matrix Address of Dependent Clause	
9	Status of Clause in Sentence	1 Single main clause
		2 First main clause of coordination
		3 Main clause medially in coordination
		4 Last main clause of coordination
		5 Dependent clause before its superordinate
		6 Dependent clause inside its superordinate
		7 Dependent clause after its superordinate (but not sentence-final)
		8 Dependent clause standing last in a sentence

10	Clausal Construction	
		1 Simple clause, not coordinated
		2 Simple, initial part in coordination
		3 Simple, medial part in coordination
		4 Simple, final part in coordination
		5 Complex clause, not coordinated
		6 Complex, initial part in coordination
		7 Complex, medial part in coordination
		8 Complex, final part in coordination
11	Clause Type	1 Main clause introduced by coordinator
		2 Embedded main clause
		3 Main clause, other than 1 & 2
		4 att-clause not introduced by preposition
		5 Indirect interrogative clause
		6 att-clause introduced by preposition
		7 Subject-modifying relative clause
		8 Object-modifying relative clause
		9 Relative clause modifying other elements in its superordinate clause
		A Adverbial clause
		B Comparative clause
		C Clausal contractions
12	Surface Mood of Clause	
		1 Affirmative
		2 Interrogative
		3 Imperative
		4 Exclamatory
13	Clause Structure	
		1 Active transitive
		2 Active intransitive
		3 Passive with full subject
		4 Existential
		5 Passive with formal subject only
		6 Other clauses with formal subject
		7 Fragments
14	Surface Word Order of Clause	
		1 SVX

2 SXV
3 VSX
4 VXS
5 XSV
6 XVS
7 SV
8 VS
9 V

15 Verb Complements

1 Object (+ Adv)^n
2 Complement (+ Adv)^n
3 (Adv)^n

Constituent Variables.

16 Verb Structure

1 Complex form of phrasal verb in verb chain + sig
2 Complex form of phrasal verb in chain
3 Complex form of V + sig in verb chain
4 Complex verb form in verb chain
5 Phrasal verb in verb chain + sig
6 Phrasal verb in verb chain
7 V + sig in verb chain
8 Simple verb form in verb chain
9 Complex form of phrasal verb + sig
A Complex form of phrasal verb
B Complex form of V + sig
C Complex verb form
D Phrasal verb + sig
E Phrasal verb
F V + sig
6 Simple verb form

17 Subject Structure

1 Coordination
2 NP + clause
3 Clause
4 Infinitive

5 NP + postmodifier
6 Preposed participle + NP
7 Other premodifier + NP
8 One word (and possibly an article)
9 Formal subject
A Subject gapped

18 Subject Form

1 Indefinite article
2 No article
3 Indefinite quantifier
4 Definite article
5 Proper name
6 Genitive attribute
7 Definite quantifier
8 Definite pronoun
9 Irrelevant

19 Givenness of Subject

1 Just mentioned
2 Recently mentioned
3 Previously mentioned
4 Indexically given
5 Pragmatically given
6 New
7 Irrelevant

20 Type of Coreference of Subject

1 Anaphoric pronoun
2 Cataphoric pronoun
3 Exophoric pronoun
4 Repetition
5 Synonym; paraphrase
6 Irrelevant

21 Number of Words in Subject Phrase

22 Object Structure (cf. 17)

23 Object Form (cf. 18)

24 Givenness of Object (cf. 19)

25 Type of Coreference of Object (cf. 20)

26 Number of Words in Object Phrase

27 Complement Structure

1 Coordination
2 NP + clause
3 Clause
4 Infinitive
5 NP + postmodifier
6 Preposed participle + NP
7 Other premodifier + NP
8 One word (and possibly an article)
9 Adjective + clause
A Adjective + adjective coordination
B Adj + modifier(s)
C One word adjective

28 Complement Type

1 Obligatory subject complement
2 Optional subject complement
3 Obligatory object complement
4 Optional object complement

29 Number of Words in Complement

30 Structure of Initial Adverbial

1 NP/PrepP + clause
2 (Prep+) Clause
3 Clausal contraction
4 (Prep+) infinitive
5 Prepositional phrase
6 NP or Adv Phrase
7 One word

31 Type of Initial Adverbial

1 Obligatory adverbial
2 Frame adverbial
3 Commentatory adverbial
4 Negative particle
5 Special clausal adverbial
6 Other optional adverbial

32 Semantics of Initial Adverbial

1 Time
2 Place

3 Manner
4 Degree; measure
5 Purpose; goal; condition; cause
6 Instrumental
7 Agent
8 Irrelevant

33 Givenness of Initial Adverbial (cf. 19)
34 Type of Coreference of Initial Adverbial (cf. 20)
35 Structure of Central Field Adverbial (cf. 30)
36 Type of Central Field Adverbial (cf. 31)
37 Semantics of Central Field Adverbial (cf. 32)
38 Givenness of Central Field Adverbial (cf. 19)
39 Type of Coreference of Central Field Adv (cf. 20)
40 Structure of Last but One Adverbial (cf. 30)
41 Type of Last but One Adverbial (cf. 31)
42 Semantics of Last but One Adverbial (cf. 32)
43 Givenness of Last but One Adverbial (cf. 19)
44 Type of Coreference of Last but One Adv (cf. 20)
45 Structure of Final Adverbial (cf. 30)
46 Type of Final Adverbial (cf. 31)
47 Semantics of Final Adverbial (cf. 32)
48 Givenness of Final Adverbial (cf. 19)
49 Type of Coreference of Final Adverbial (cf. 20)
50 Special Clausal Adverbial inside a Functional Element

1 In subject
2 In object
3 In complement
4 In other adverbial

51 Number of Adverbials in the Clause

Transformations.

52 Type of Topicalized Constituent

1 Object
2 Object as quotation
3 Adverbial from central field position
4 Adverbial from end-field position

5 Complement
6 Subject in existential clause
7 VP
8 WH question-word
9 Relative pronoun

53 Subject Postponement to the End-field
1 NP subject in existential construction
2 NP Subject other than 1
3 Clause as subject
4 Infinitive as subject

54 Extraposition and Quantifier Movement
1 Extraposition of part of subject
2 Extraposition of part of object
3 Extraposition of part of complement
4 Extraposition of part of adverbial
5 Extraposition of functional element from clause
6 Movement of quantifier
7 Movement of head, quantifier remains

55 Type of Cleft Construction
1 Subject clefted
2 Object clefted
3 Complement clefted
4 Adverbial clefted
5 Pseudo-cleft

56 Left Dislocation
1 Subject
2 Object
3 Complement
4 Initial adverbial
5 Other adverbial

57 Right Dislocation (cf. 56)

58 Raising
1 Subject
2 Object -- generic
3 Object -- non-generic
4 Adverbial

59 Passive

1 -s-passive without agent
2 -s-passive: agent as initial adverbial
3 -s-passive: agent as central field adverbial
4 -s-passive: agent as last but one adverbial
5 -s-passive: agent as final adverbial
6 bli passive without agent
7 bli passive: agent as initial adverbial
8 bli passive: agent as central field adverbial
9 bli passive: agent as last but one adverbial
A bli passive: agent as final adverbial
B man passive

60 Verb Phrase Deletion

1 Forward, outer
2 Forward, inner
3 Backward, outer
4 Backward, inner

61 Ellipsis of Functional Elements (Gapping)

1 Subject
2 Object
3 Adverbial
4 Verb
5 S + O
6 S + V
7 V + O
8 V + A

62 Ellipsis of Part of the Verb

1 Auxiliary in complex form
2 göra
3 fara

63 Ellipsis of the Head of a Noun Phrase

1 First part of coordinated NPs
2 Second part of coordinated NPs

3 First part of NPs not in coordination

4 Second part of NPs not in coordination

64 Number of Transformations in the Clause

REFERENCES

Andersson, Erik, (1976) Verbfrasdeletion i svenskan. Nysvenska Studier 55-56, 47-76.

_____. (1977) Verbfrasens struktur i svenskan. Publications of the Research Institute of the Abo Akademi Foundation, Nr. 18, Abo.

Diderichsen, Paul. (1946) Elementaer dansk grammatik. Gyldendal, København.

Hakulinen, Auli & Fred Karlsson. (forthcoming) Näkökulmia nykysuomen lauseoppiin. SKS, Helsinki.

Kohonen, Viljo & Jussi Salmela. (this volume) "Aineiston valinnan ja automaattisen tietojenkäsittelyn ongelmia kielitieteellisessä tutkimuksessa"

Lahti, Marjatta. (1978) Faktorer som styr fundamenteringen i Vilhelm Mobergs roman "Din stund på jorden". Pro gradu thesis, Abo Akademi.

MATERIAL

Carpelan, Bo. (1968) Bågen. Bonnier, Stockholm.

Lindgren, Astrid. (1973) Bröderna Lejonhjärta. Rabén & Sjögren, Stockholm.